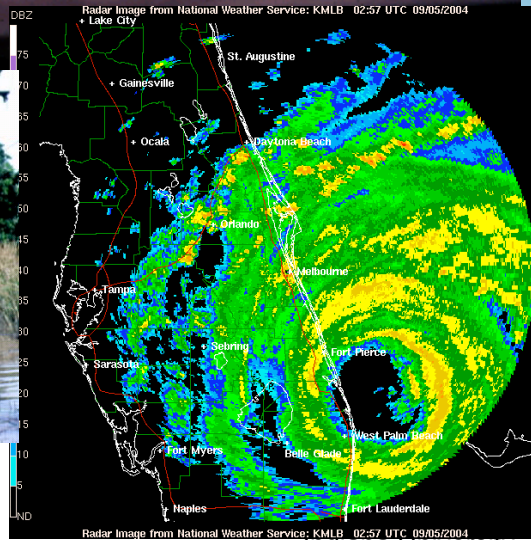


# TrustCell: Towards End-to-End Trustworthiness in Data-Driven Scientific Computing

The slide features five yellow circles of varying sizes and styles. Two are solid yellow, one is a thin yellow outline, and two are semi-transparent yellow. They are arranged in a decorative pattern around the title and authors.

Sangmi Lee Pallickara and Beth Plale  
Computer Science Department,  
Indiana University

# Context: dynamic mesoscale weather forecasting and broadening engagement



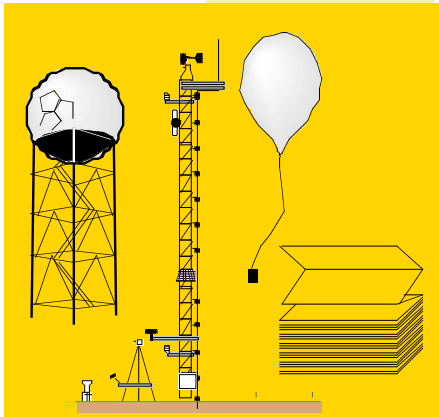
© 1993 Roger Edwards

© Warren Fa



Indiana University  
WGSA 2006

# Traditional Methodology



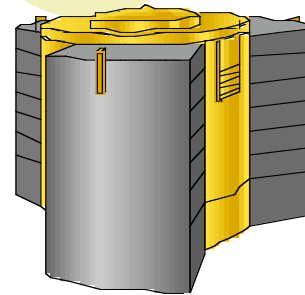
## STATIC OBSERVATIONS

Radar Data  
Mobile Mesonets  
Surface Observations  
Upper-Air Balloons  
Commercial Aircraft  
Geostationary and Polar Orbiting Satellite  
Wind Profilers  
GPS Satellites



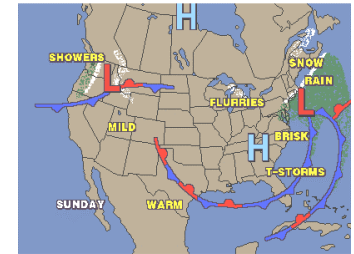
## Analysis/Assimilation

Quality Control  
Retrieval of Unobserved Quantities  
Creation of Gridded Fields



## Prediction/Detection

PCs to Teraflop Systems



## Product Generation,

Display,  
Dissemination

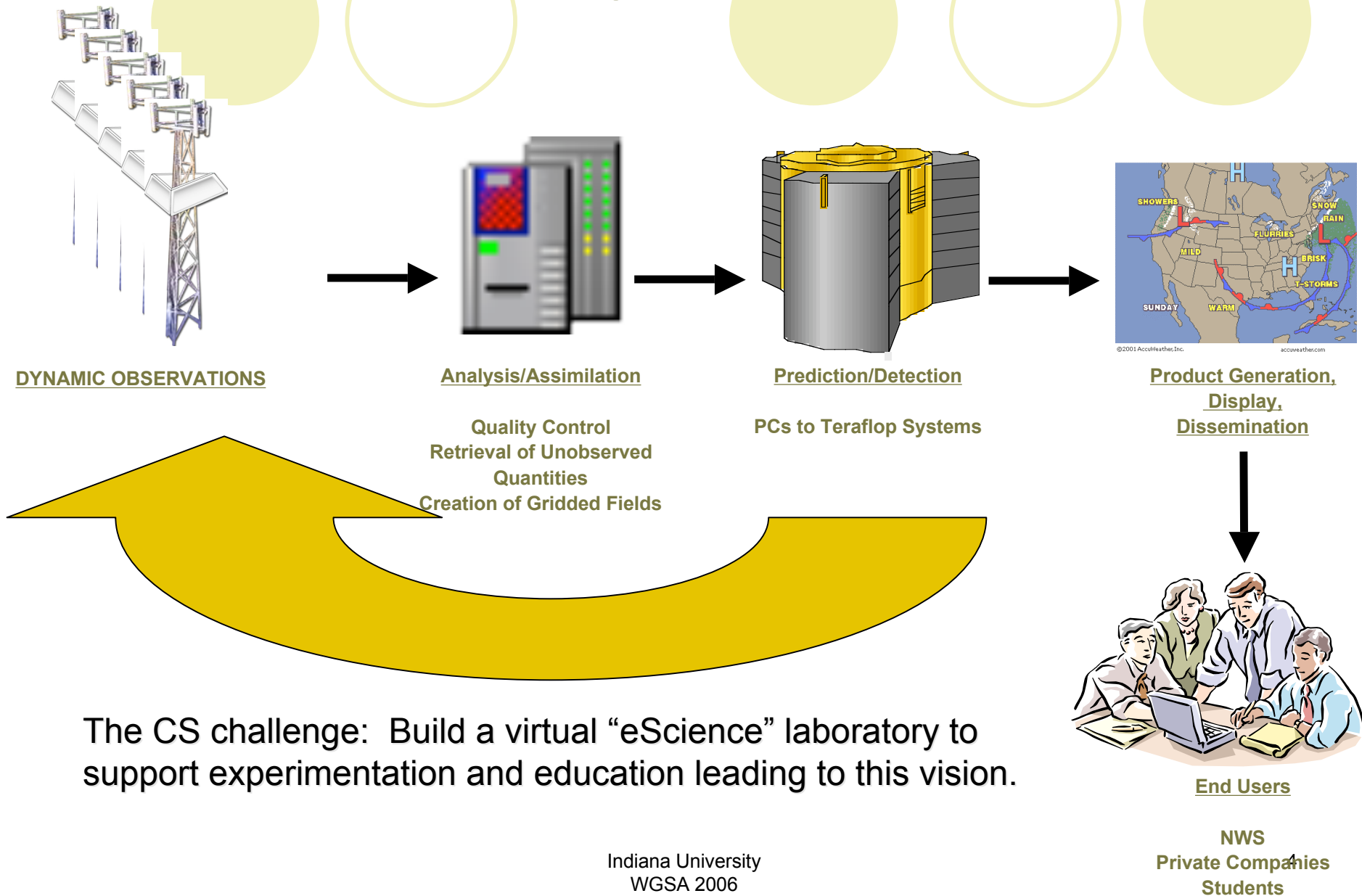


## End Users

NWS  
Private Companies  
Students


Process is entirely serial  
and static (pre-scheduled):  
No Response to the Weather!


# The LEAD Vision: A Paradigm Shift



# LEAD Science Gateway

<http://portal.leadproject.org>

**LEADPORTAL**  
LINKED ENVIRONMENTS FOR ATMOSPHERIC DISCOVERY

SPONSORED BY THE NATIONAL SCIENCE FOUNDATION 

HOMEABOUT LEADDATA SEARCHVISUALIZEEDUCATIONRESOURCESHELP

WelcomeCreate AccountForgot your password?

## > WELCOME TO THE LEAD PORTAL



Linked Environments for Atmospheric Discovery (LEAD) makes meteorological data, forecast models, and analysis and visualization tools available to anyone who wants to interactively explore the weather as it evolves. The LEAD Portal brings together all the necessary resources at one convenient access point ... [read more](#)

LOGIN

☐ Remember my login

[Forgot your password?](#)[Create new account](#)

## > FEATURES FOR ANYONE INTERESTED IN THE WEATHER

<b>Researchers</b>	With university, government, or industry affiliations	GET FEATURES
<b>Educators</b>	At college and university level, high school, or middle schools	GET FEATURES
<b>Students</b>	At graduate, undergraduate, middle and high school levels	GET FEATURES
<b>Visitors</b>	Newcomers and the curious	GET FEATURES

### QUICK LINKS

- [Live Weather](#)
- [LEAD Grid](#)
- [Glossary](#)
- [Website Help](#)
- [Frequently Asked Questions](#)

## > POPULAR TOOLS

Visualize Weather Data

**Integrated Data Viewer** | MORE >




Make a Forecast or Analysis

**Experiment Builder** | MORE >



Access Weather Data

**Geographic Region Search** | MORE >

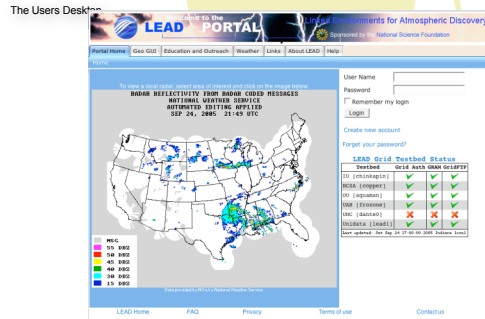


### THE LEAD TEAM

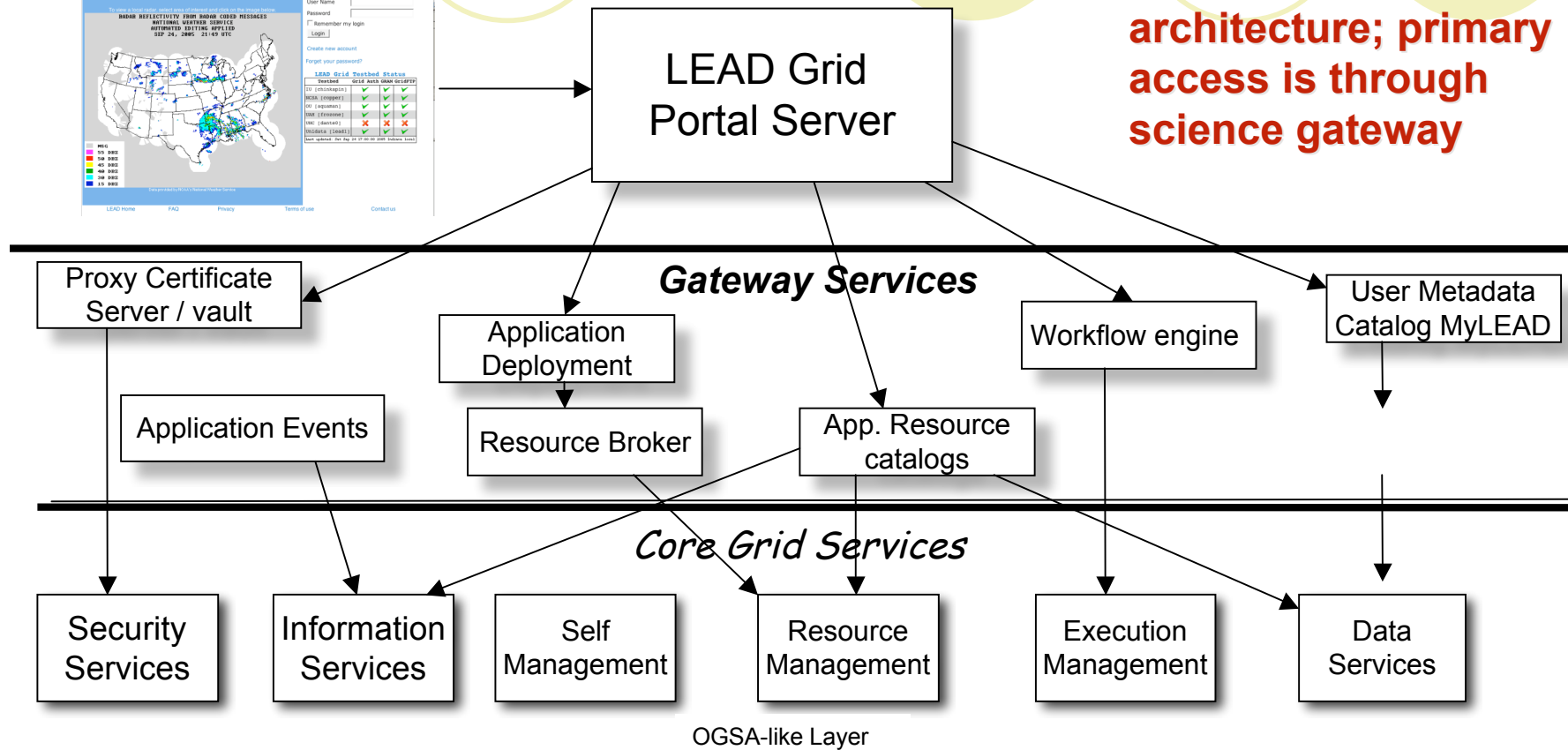




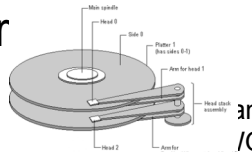
# Architecture



**Grid/web Service architecture; primary access is through science gateway**

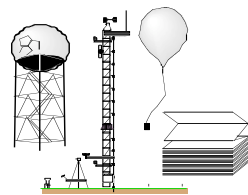
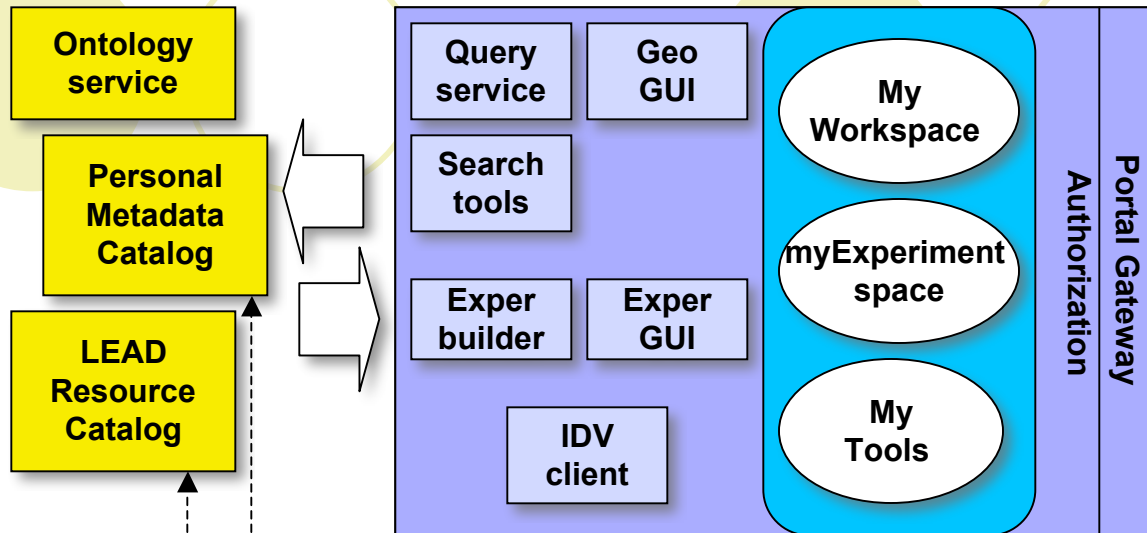


## Physical Resource Layer

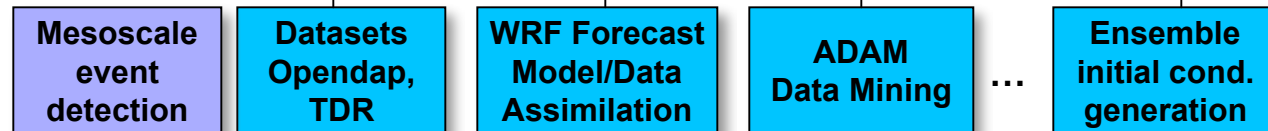


# Flow

User  
Workspace  
and  
Resources



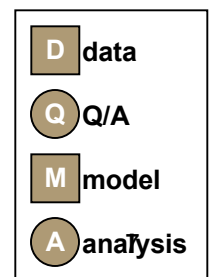
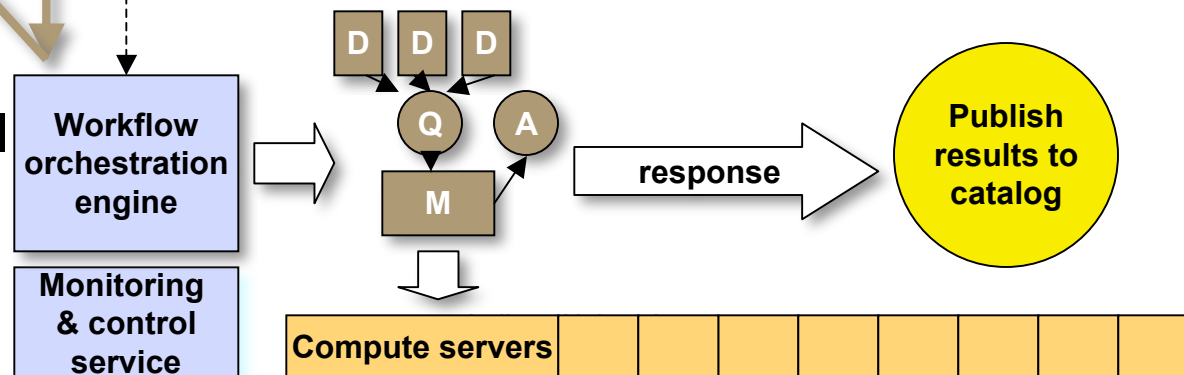
Registered with catalog



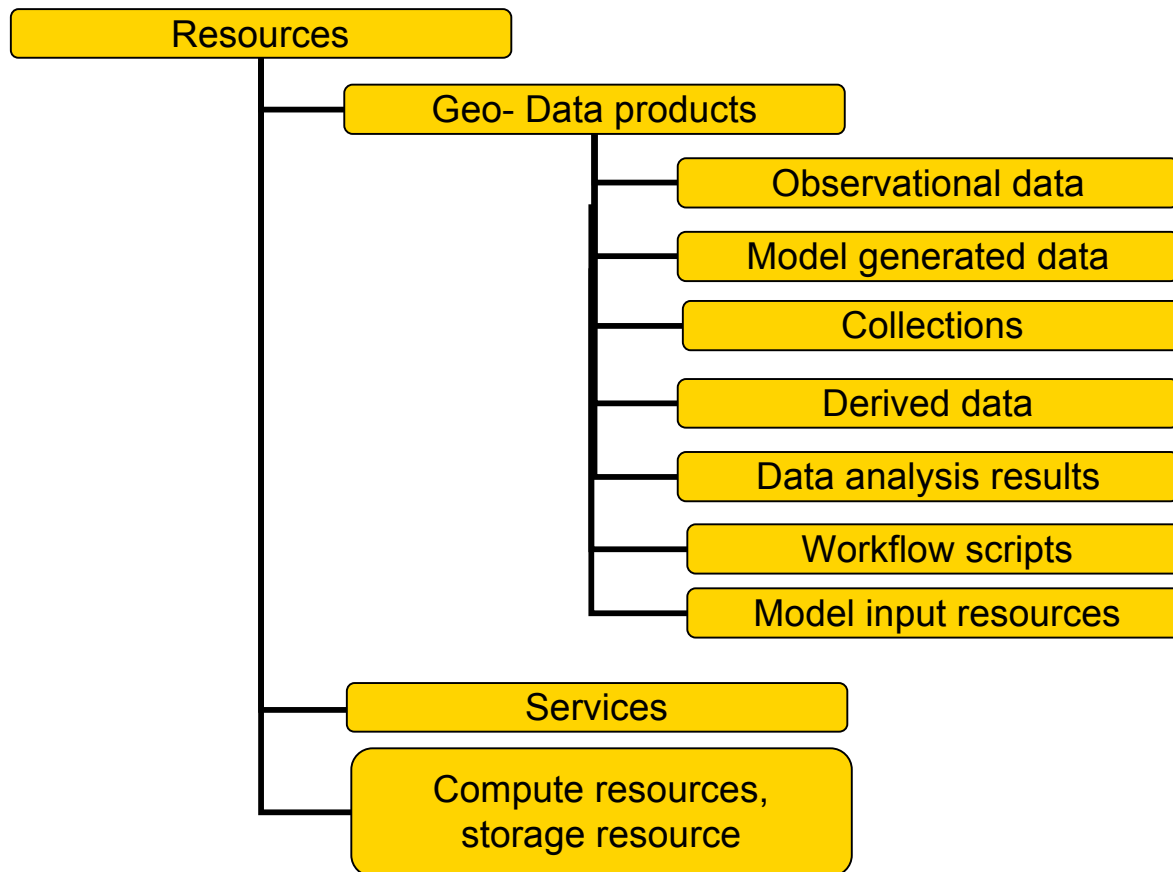
Middleware  
Services  
and  
Resources

Used in experiment

Computational  
Layer



# Data Subsystem: Management of Multiple Product Categories



## Personal data products

-- user's experiment products, personal collections, scripts, input config params.

*Scott's talk*

## Community data products

-- data, mostly observational, made available to LEAD virtual community

## External products

-- new and future data sources.



# Data collections increasingly crucial to exploratory research and education in science and engineering

- *Current influential technology factors:*
  - *Powerful and affordable sensors, processors, instruments, automated equipment*
  - *Reductions in storage costs make cost-effective to maintain large data collections*
  - *Internet makes it easier to share data*
- *As result, researchers increasingly carry out computational science investigations using data generated by others and widely distributed.*
  - *Research in biodiversity and ecosystems, global climate change, meteorology, space science depend on ability to combine vast quantities of digital information with complex models and analytical tools.*

# Problem Domain: secure storage, retrieval, and access to scientific data collections

- Digital data collections\* are the foundation for analysis using *automated analytical tools*
- Long-lived data undergoes constant re-analysis for
  - improved algorithms or
  - with alternate use in mind.
- Analysis depends not just on sensed or computer-generated data but on the *metadata* that characterizes the environment and the sensing instrument.

\*Data - text, numbers, images, video or movie clips, audio, software, algorithms, equations, models, simulations

\*Digital data collections - data itself, and infrastructure, organizations needed to preserve access to the data.

# Petascade data collections require new work style

- Analysis tools growing more complex
  - Many analysis algorithms are super-linear, often needing  $N^2$  or  $N^3$  time to process  $N$  data points
- I/O bandwidth has not kept pace with storage capacity
  - Capacity increase 100-fold while storage bandwidth increase 10-fold
- Too many files (> 1million) for a local file system to manage
  - File name and directory hierarchy not enough
- Can't download dataset to laptop and process, analyze, visualize
- **Therefore, move end-user's program to the data, only communicate questions and answers**

# Problem Statement

- Users need to be able to park data on their project's grid and know it is protected.
  - Services accessing the repository do so only on my behalf and not someone else's.
- Service-oriented middleware through proxy delegation is well suited to carrying out actions on behalf of a user.
- However, SOA middleware does not address the user's privacy to the very end of the remote resource chain, where resources include accessing data collections.
- End-to-End trustworthiness is required!

# Our Approach



- We propose a Trust model that represents the trust relationship between the users and their remote resources in the Grid system.
  - Over different security schemes.
  - Over resources used by multiple Grid applications.
  - Over the resources that themselves contain nested services.

# The Concept of “trust”

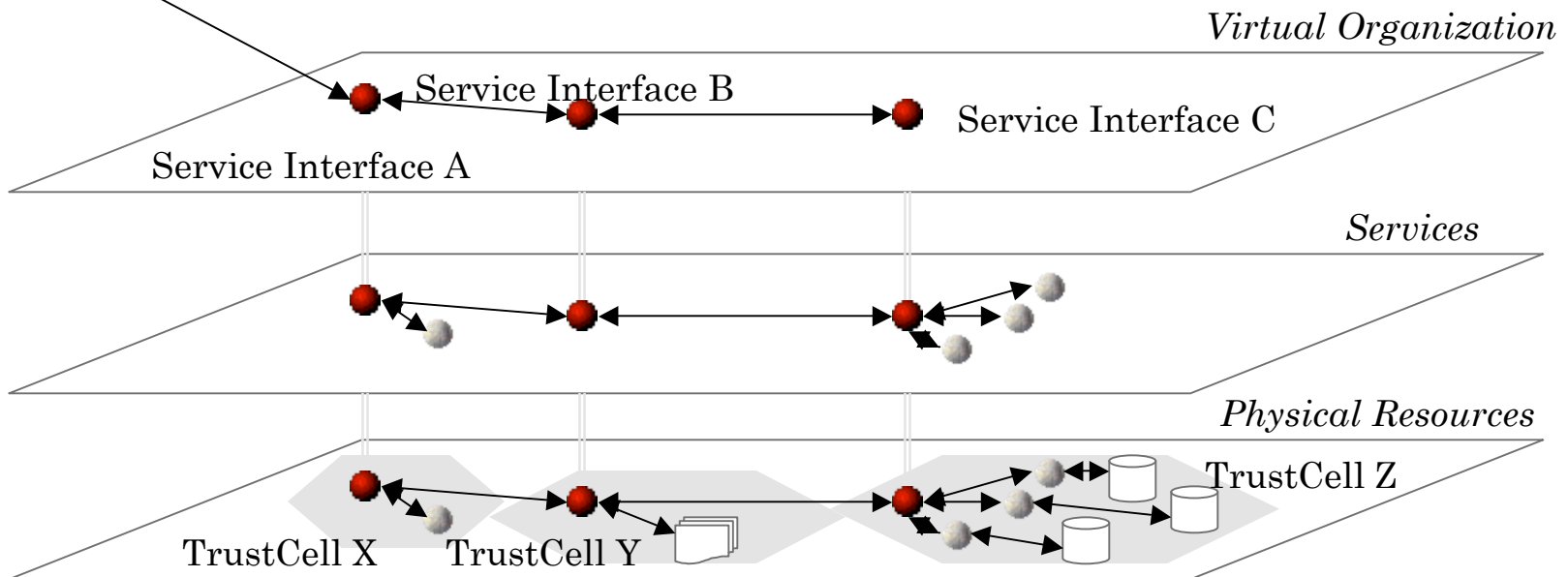
- Trust is a subjective quality which individuals place on one another based on a form of cost-benefit analysis.
  - *If an individual is confronted with an ambiguous path, this path can lead to an event that can be perceived to be beneficial ( $Va^+$ ) or to an event perceived to be harmful ( $Va^-$ )*
  - *The occurrence of ( $Va^+$ ) or ( $Va^-$ ) is contingent on behavior of another person; and*
  - *The strength of ( $Va^-$ ) is greater than the strength of ( $Va^+$ ).*
  - *If he chooses to take such a path, he makes a trusting choice*  
*Deutsh, 1962.*
- In other words: Sees path. Outcome could be good or bad, bad is stronger, and outcome depends on another person. If takes path, then trusts person.



# Starting from the Observation of the Grid Services



“Should I trust?”

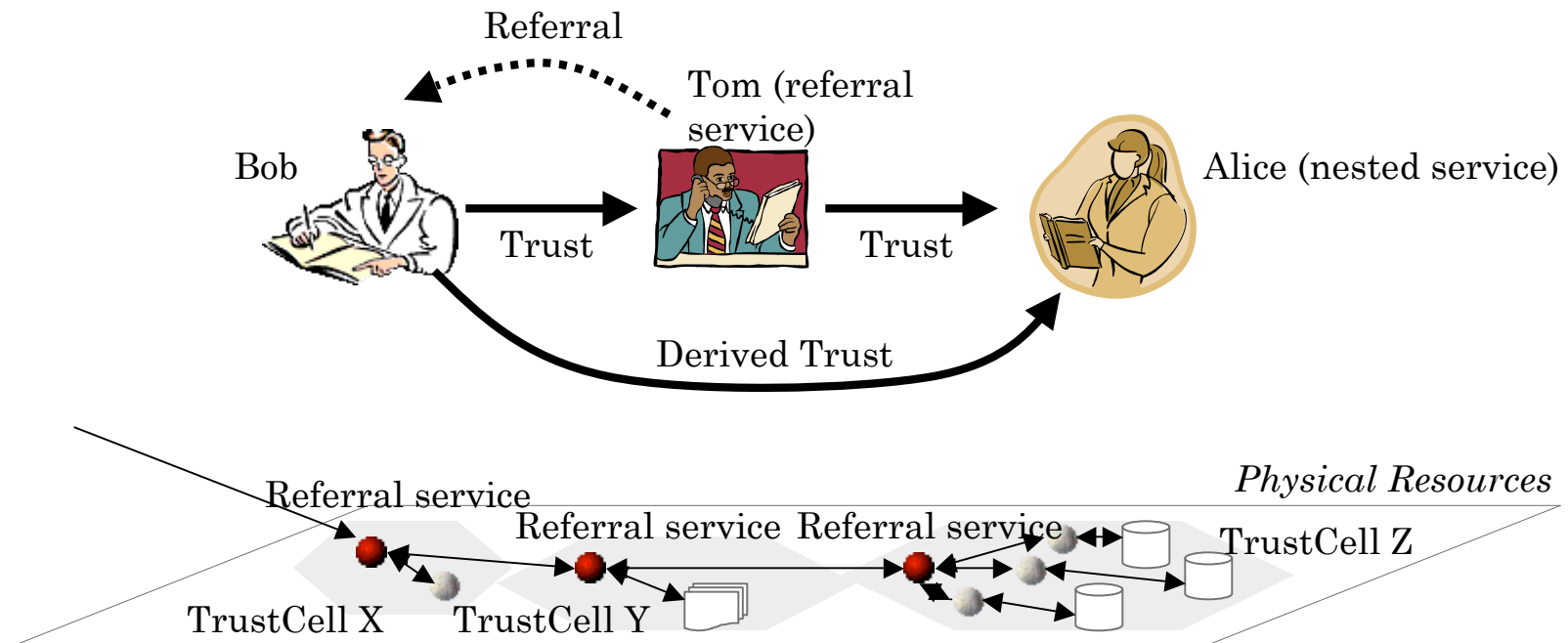


# Trust Cells in a Grid Environments

- Trust cell is the minimum unit wherein the trustworthy relationship between resources is ensured.
- Trust cell is managed by single security policy.
- A organization can have one or more trust cell, and a trust cell can be composed by the resources from one or more organizations
- A trust cell MUST provide at least one *referring service* (top level service) to the computational community.

# Referring Service in a Trust Cell

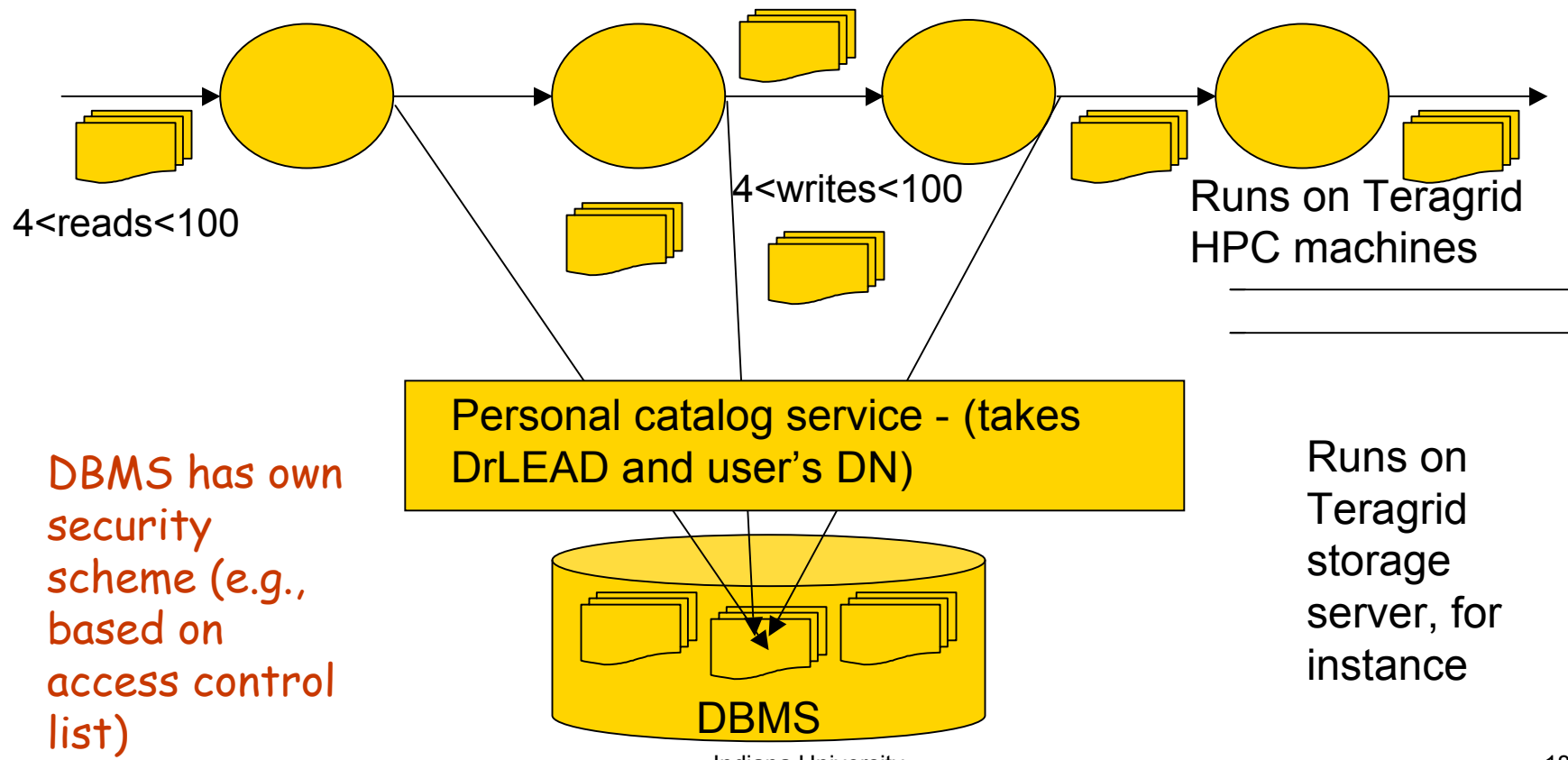
- Referring service provides trustworthy recommendation to the user for their remote resources.
- In TrustCell model, the referral service is a reliable recommender about the resources of the TrustCell.



# Automated Metadata and Data Collection

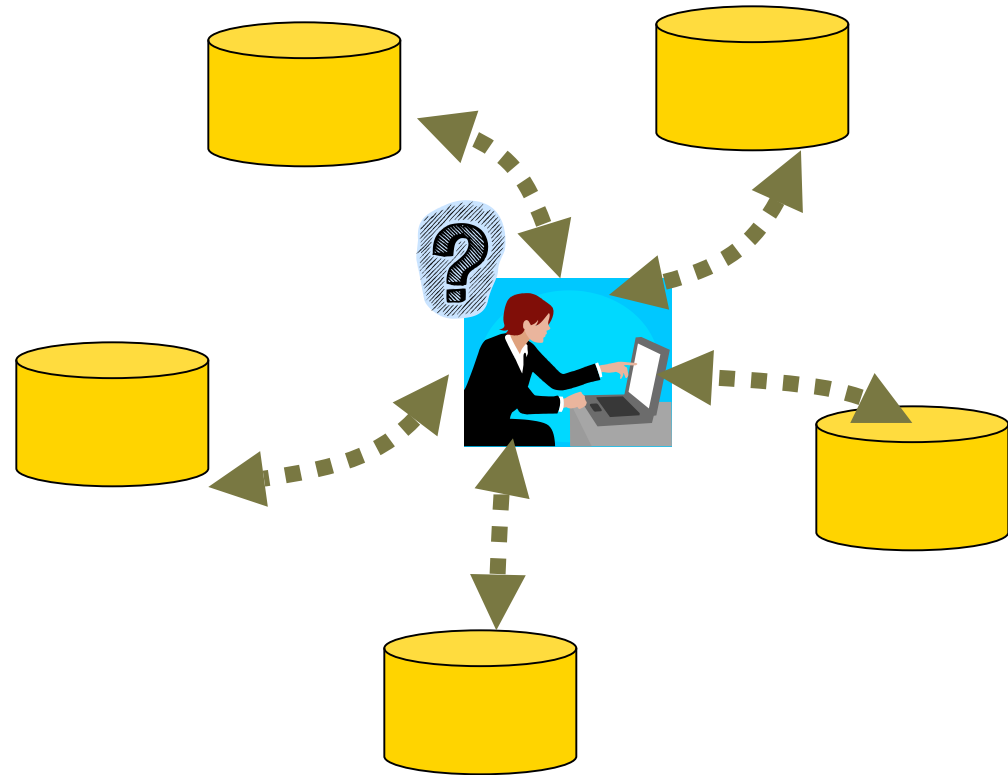
Provides:

- 1) Automated metadata collection
- 2) Automated structure and organization



# Application of TrustCell to Personal Workspace

- Access to the personal data located on a replicable services
- Select most trustworthy MyLEAD server under user's context
- Get reference from each site about “final mile” access to data



# Requirements of the Referring Services

- A local and global trusted delegator in the predefined context.
- Able to verify the competence of a requester in order to delegate him/her to work in the context.
- Contains revocation power over the released delegation certificates.
- Provides internal trustworthiness within the TrustCell.



# Ability of Trust Cell model to model alternate security scheme

- Trust cell can be implemented over the existing Grid Security Infrastructure (GSI) for the fundamental identity authentication and delegation.
- Here we present how the fine grained user capability service is modeled within the Trust Cell model.

# Ongoing work: Trust Server

- Trust Referring service generates trust factors based on their in-house security scheme.
- Users or services should often select a resources among the replications of distributed community resources.
- Based on the Trust Cell model, Trust server provides generalized reputation of the replicated resources.

# Future Works: Trust Server

- Trust server provides and updates the value of the trustworthiness of the trust cell.
- Trustworthiness is calculated based on the information from the referring service of the trust cell.
- Trustworthiness considers multiple factors such as, satisfaction of the service, security requirements, context specific policies, and the credibility of the referring service.
- Trustworthiness is updated based on the given time windows of activities.

# Related Approaches

- Identity trust models: KeyNote, PolicyMaker, Simple Public Key Infrastructure (SPKI), Simple Distributed Security Infrastructure (SDSI)
  - Manage security in large-scale distributed networks through credentials that delegate permissions
  - These approaches investigate trust management between service interfaces without consideration for the resources these service interfaces themselves access.

## Related work



- Reputation-based trust model: XREP, P-Grid Trust Model, NICE Trust Inference Model
- Social Networks-based Trust Model: Community-based Reputation, Regret, NodeRanking

# Conclusion



- The TrustCell model extends the inter-service security infrastructure to the end-to-end trusted relationship between user and physical resources.
- TrustCell: Minimum unit of resources assuring trustworthy relationship between them.
- Referring service: Service provides trust-related information of the TrustCell to outside of the TrustCell.
- TrustCell model can adapt from current security models.



Thanks to ...



- Collaborators

- Dennis Gannon, IU

- Oklahoma Univ, Unidata, NCSA, UNC, UAH

- DDE lab members

- Sangmi Lee, Scott Jensen, Yiming Sun, Ning Liu in particular

- NSF EIA-0202048, NSF ATM-0331480